

Amendments to the Specification:

Please replace paragraph [0001] with the following amended paragraph:

- 5 **[0001]** This Application incorporates by reference:
U.S. Patent Application Serial No. 09/609,325, entitled "System and Method for
Teaching Expository Writing Using Microanalysis of Text" by L. POLANYI et al.;
U.S. Patent Application Serial No. 09/630,731, entitled "A System and Method for
Generating Text Summaries" by L. POLANYI et al.;
10 U.S. Patent Application Serial No. 09/883,345, entitled "System and Method for
Generating Analytic Summaries" by L. POLANYI et al.;
U.S. Patent Application Serial No. 10/231,732, entitled "A System and Method for
Summarization Containing Combining Natural Language Generation with Structural
Analysis" by L. POLANYI et al.; and
15 U.S. Patent Application Serial No. 10/435,036, entitled "Systems and Methods for
Grammatical Text Condensation" by S. RIEZLER et al. each, in their entirety.

Please replace paragraph [0003] with the following amended paragraph:

- 20 **[0003]** Corporations, government regulators, intelligence analysts and other
researchers and knowledge workers are faced with reviewing and understanding an
ever-increasing volumes-volume of information. Some researchers have attempted to
reduce the volume of information by creating summaries based on the most relevant
information. Relevance measures are frequently based on non-structural features such
as keywords and/or term-frequency and inverse document frequency statistics for sets
25 of documents in an information repository. However, these conventional, non-
structural text summaries are frequently disjointed and difficult to read. Moreover,
elements selected for these non-structural text summaries are based on statistical
measures of importance. Therefore, pronouns in sequentially selected text summary
elements may refer to elided preceding references reducing the usefulness of these
30 conventional summaries.

Please replace paragraph [0007] with the following amended paragraph:

[0007] Fig. 1 is an overview of an exemplary hybrid summarization system 100 according to this invention. The hybrid summarization system 100 is connected via communications link 99 to a web-enabled personal computer 300, a web-enabled tablet computer 400 and an information repository 200 containing documents 1000-1002_10000-10002.

5

Please replace paragraph [0008] with the following amended paragraph:

[0008] In one of the various exemplary embodiments according to this invention, a user of the web-enabled personal computer 300 initiates a request for a hybrid text summary of document 1000_10000 contained in the information repository 200. The request is forwarded over communications link 99 to the hybrid summarization system 100. The hybrid summarization system 100 mediates the request for the hybrid text summary. The hybrid summarization system 100 then retrieves the document 1000_10000 from the information repository 200. The text of document 1000_10000 is then segmented into discourse constituents based on the determined segmentation rules for the theory of discourse analysis.

10

15

Please replace paragraph [0009] with the following amended paragraph:

[0009] In various exemplary embodiments according to this invention, the segmentation rules are used to segment the text into basic discourse constituents units (BDUs). The basic discourse constituent units are determined based on segmentation rules developed for the Unified Linguistic Discourse Model (ULDM), the Rhetorical Structure Theory (RST), the Discourse Structure Theory, the Structured Discourse Representation Theory (SDRT) or any other known or later developed theory of discourse analysis. For example, each basic discourse constituent in the Linguistic Discourse Model is composed of a linguistic gesture interpreted relative to context that communicates: 1) information about not more than one state of affairs in a possible world; and 2) function information that encodes information about how previously occurring or possibly subsequent linguistic gestures relate structurally, semantically, interactively or rhetorically to other units in the discourse or context in which the discourse takes place. The sentential syntactic reflex of a linguistically realized basic discourse constituent unit is a segment accessible for continuation from outside the

20

25

30

sentence in which it is found. Reference to an event is necessary but not sufficient for a word to be a head of most discourse ~~constituent~~constituents. Thus, auxiliary and modal verbs will not form separate discourse constituent from their main verbs. ~~Since verbs, since~~ they do not refer to separate events. Other modal constructions that involve infinitives such as “have to”, “ought to”, and the like also constitute a single segment with their complements by the same reasoning.

5 Please replace paragraph [0010] with the following amended paragraph:

[0010] Thus, to determine discourse constituent units, a candidate discourse constituent segment must reference an event and provide the context sufficient to “continue” the event later in the discourse. Equi verbs such as “try” and “persuade” and raising verbs such as “seem” and “believe” form separate arguments from their verbal complements since both events can be ~~continuescontinued~~. In contrast, even though event nominals, including gerunds refer to events possibly distinct from the 10 verbs they are arguments or adjuncts of, those events are not easily ~~continuescontinued~~.

15 Please replace paragraph [0013] with the following amended paragraph:

[0013] In another exemplary embodiment according to this invention, the user of web-enabled tablet computer 400 requests a hybrid text summary of the document ~~100110001~~ in the information repository 200. The request is forwarded over communications link 99 to the hybrid summarization system 100. The hybrid summarization system 100 retrieves the document ~~100110001~~ from the information repository 200 over communications link 99. A structural representation of discourse 20 is determined for the requested document ~~100110001~~ based on a theory of discourse analysis that specifies the links between determined discourse constituents. Statistical analysis, keywords, knowledge base searches or any known or later developed method of determining important information may be used to determine important discourse 25 constituents in the text. For example, statistical and/or keyword based techniques are applied to determine discourse constituents having relevance scores which exceed a threshold importance score. The determined important discourse constituents and 30

supporting preceding discourse constituent within the structural representation of discourse define a sub-tree of discourse constituents.

Please replace paragraph [0022] with the following amended paragraph:

5 **[0022]** Structural rules of the Unified Linguistic Discourse Model specify the type of relation that holds between a structural representation of discourse and a discourse constituent being conjoined to that representation. The structural rules constrain the structural attachment point of the discourse constituent to the structural representation as well as the nature of that attachment. The discourse constituent
10 being conjoined to the structural representation will be made up of one or more units derived from a single sentence and conjoined on the basis of the unit corresponding to the main clause of the discourse unit. Thus, it will be apparent that structural rules are useful in reducing the number of candidate attachment points that ~~need s~~s ~~needs~~ to be considered when adding discourse constituents to the structural representation of
15 discourse.

Please replace paragraph [0030] with the following amended paragraph:

20 **[0030]** Fig. 3 is a first hybrid summarization system according to an exemplary embodiment of this invention. The hybrid text summarization system 100 is comprised of a memory 20; a processor 30; a discourse constituent segmentation routine or circuit 40; a discourse constituent link determination routine or circuit 50; a relevance score determination routine or circuit 60; and a linked preceding discourse constituent determination routine or circuit 70, each connected to input/output circuit 10 via the communications link 99, and to a web-enabled personal computer 300, a
25 web-enabled tablet computer 400 and an information repository 200 containing documents ~~1000-1002~~10000-10002.

Please replace paragraph [0031] with the following amended paragraph:

30 **[0031]** In one of the various exemplary embodiments according to this invention, a user of the web-enabled personal computer 300 initiates a request for a hybrid summary of document ~~1000~~10000 contained in information repository 200. The request is forwarded over the communications link 99 to the input/output circuit

10 of the hybrid summarization system 100. The processor 30 activates the input/output circuit 10 to retrieve the requested document ~~100010000~~ from the information repository 200 over the communication link 99 and store the document ~~100010000~~ in memory 20.

5

Please replace paragraph [0047] with the following amended paragraph:

[0047] Fig. 5 is a second hybrid summarization system according to an exemplary embodiment of this invention. The hybrid summarization system ~~100110~~ is comprised of a memory 20; a processor 30; a segmentation rule storage 40; an attachment rule storage 50; a structural representation of discourse routine or circuit 60; a relevance score routine or circuit 70; a percolation routine or circuit 80; and a display routine or circuit 90, each connected to the input/output circuit 10; and via communications link 99 to a web-enabled personal computer 300, a web-enabled tablet computer 400, and an information repository 200 containing documents ~~1000-100210000-10002~~.

10
15

Please replace paragraph [0048] with the following amended paragraph:

[0048] In one of the various exemplary embodiments according to this invention, a user of the web-enabled personal computer 300 initiates a request for a hybrid text summary of document ~~100010000~~ contained in the information repository 200. The request is forwarded over communications link 99 to the input/output circuit 10 of the hybrid summarization system ~~100110~~. The processor 30 activates the input/output circuit 10 to retrieve the document ~~100010000~~ from the information repository 200 over the communication link 99 and store it in memory 20.

25

Please replace paragraph [0073] with the following amended paragraph:

[0073] The pattern portion 630 contains the value “(Discrete_Event and (PostNominal_Modifier or Relative_Clause or Participial_Modifier))-or (~~NOT(Discrete_Event) and Prepositional_Phrase_Modifier~~)”. This indicates that these constraints must be determined to be associated with the candidate segment before the rule will fire.

30

Please replace paragraph [0087] with the following amended paragraph:

5 **[0087]** The fourteenth row of the exemplary data structure for storing segment rule information contains the value “14” in the identifier portion 610 and the value “Initial_Modifier” in the type portion 620. This indicates that the fourteenth fourteenth rule is an initial modifier type of rule. The pattern portion 630 contains the value “(Initial_Modifier and Scene_Setting)”. This indicates that the candidate segment must be an initial modifier and be associated with scene setting to trigger a firing of the rule.

10 Please replace paragraph [0090] with the following amended paragraph:

15 **[0090]** The first row of the exemplary data structure for storing attachment rule information contains “1” in the rule identifier portion 710. The rule portion 720 contains the compound rule ‘<Sx: (M-BDU/Type) = “BetweenParenthesis”; St: AP is BottomOfTree;> → Subordinate(AP, M-BDU)’. Rule “1” is a compound rule comprised of a first syntactic rule_constraint denoted by ‘Sx: (M- BDU/Type) = “BetweenParenthesis”’ indicating the main BDU is delimited by parentheses, and a second structural constraint “St: AP is BottomOfTree” indicating that the attachment point must be at the bottom of the open right tree of the structural representation of discourse. The first and second constraints are combined between the open constraints “<” and close constraints “>” tags indicating that both constraints must match to trigger the subordination of the main basic discourse unit (M-BDU) to the current attachment point indicated by the attachment point variable (AP).

20 Please replace paragraph [0091] with the following amended paragraph:

25 **[0091]** The second row of the exemplary data structure for storing attachment rule information contains “2” in the rule identifier portion 710. The rule portion 720 contains the compound rule ‘<Sx: (AP/Tense) = “Present”; Sx: (M- BDU/Tense = “Past”; > → Subordinate(AP, M-BDU)’. Rule “2” is a compound rule comprised of a first syntactic constraint denoted by ‘Sx: (AP/Tense) = “Present”;’ indicating the candidate attachment point discourse constituent must be in the present tense and a second syntactic constraint ‘Sx: (M-BDU/Tense)=”Past”’ indicating that the main basic discourse constituent to be attached to the attachment point (AP) must

be in the past tense to trigger the rule. The first and second rules-constraints are combined between the open constraint “<” and close constraint “>” tags indicating that both constraints must match to trigger the subordination of the main basic discourse unit (M-BDU) to the current attachment point (AP).

5

Please replace paragraph [0093] with the following amended paragraph:

[0093] The fourth row of the exemplary data structure for storing attachment rule information contains “4” in the rule identifier portion 710. The rule portion 720 contains the compound rule “<LS: (AP/Lexeme) hasa (M-BDU/Lexeme: \$1: \$2; Sx: (AP/*/COMP) is \$1; Sx: (M-BDU/*/SUBJ) is \$2; > → Subordinate(AP, M-BDU)”. Rule “4” is a compound rule comprised of a lexical-semantic constraint, followed by two syntacticconstraintconstraints. The first, and second and third constraints are combined between the open constraint “<” and close constraint “>” tags indicating that all three constraints both constraint must fire to trigger the subordination of the main basic discourse unit (M-BDU) to the current attachment point (AP).

10

15

Please replace paragraph [0094] with the following amended paragraph:

[0094] The fifth row of the exemplary data structure for storing attachment rule information contains “5” in the rule identifier portion 710. The rule portion 720 contains the compound rule ‘<Sx: (M-BDU/Type) = “Temporal”; St: M-BDU isSubordinatedTo AP;,> → CreateNary(Cx, AP, M-BDU)’. Rule “5” is a compound rule comprised of a first syntactic constraint denoted by ‘Sx: (M-BDU/Type) = “Temporal”’,. This indicates the main basic discourse unit (M-BDU) discourse constituent must be a “Temporal” type of discourse constituent. The second syntactic constraint “Sx: M-BDU isSubordinatedTo AP;” indicates that the main basic discourse unit M-BDU is subordinated to the attachment point. The first and second constraints are combined between the open constraint “<” and close constraint “>” tags indicating that both constraintconstraints must match to trigger the coordination of the attachment point (AP) and the main basic discourse unit (M-BDU). The term “Cx” indicates coordination subtypes such as Question/Answer subtype coordinations and the like.

20

25

30

Please replace paragraph [0096] with the following amended paragraph:

[0096] The seventh row of the exemplary data structure for storing attachment rule information contains “7” in the rule identifier portion 710. The rule portion 720 contains the compound rule ‘<Sm: Σ(AP) is “Irrealis”; Sm: Σ (M-BDU) is “Realis”; Sx: if AP then M-BDU; > → CreateNary(HYP, AP, M-BDU)’. Rule “7” is ~~requires~~requires that the AP be “Irrealis” and the main basic discourse constituent ~~“Irrealis”~~be “Realis” and the structural constraint that the attachment point (AP) match before the rule is fired to create the n-ary relationship between the attachment point and the main basic discourse constituent. Similarly, rows eight through twelve describe other Unified Linguistic Discourse Model attachment rules useful in determining a structural representation of discourse.

Please replace paragraph [0097] with the following amended paragraph:

[0097] The ninety-ninth row of the exemplary data structure for storing attachment rule information contains “99” in the rule identifier portion 710. The rule portion 720 contains the compound rule ’<St: AP is BottomOfTree; Px: Ifnot Anyrule applies;> → Subordinate(AP, M-BDU)’. Rule “7” is a compound rule comprised of a first structural constraint “St: AP is BottomOfTree indicating the candidate attachment point must occur at the bottom of the tree. A preemptor constraint “Px: IfnotAnyrule applies;” is matched only if no other ~~constraint~~constraints apply. The first and second constraints are combined between the open constraint “<” and close constraint “>” tags indicating that both ~~constraint~~s-constraints must match to trigger the subordination of the main basic discourse unit (M-BDU) to the current attachment point (AP).

25

Please replace paragraph [0103] with the following amended paragraph:

[0103] In step S2214, the relevance score of each subordinated node is assigned to the subordinating discourse constituent if the relevance score of the subordinated discourse constituent node is more relevant and the node appears to the left of the subordinated node. This ensures that supporting subordinating discourse constituent nodes are included in the hybrid text summary but further constrains the addition of supporting discourse constituents to discourse constituents on the left. In

the structural representation of the Linguistic Discourse Model, the left child node of a subordination node is the subordinating node. However, it will be apparent that due to differences in the structural representation of discourse associated with various other theories of discourse analysis, other methods of identifying the supporting preceding 5 discourse constituent information may be used without departing from the scope of this invention. Control then continues to step S2235-S2216 where for each coordination or n-ary or binary discourse constituent node, more relevant relevance scores of child discourse constituents are assigned to all child discourse constituents to the left. Control then continues to step S2218 and immediately jumps to step S2300 of 10 Fig. 4.

Please replace paragraph [0108] with the following amended paragraph:

[0108] The score of any subordinated node is assigned to the subordinating node if the subordinated score is lower in step S2270S2242. Control then continues to 15 step S2275S2244. In step S2244, the score of any coordination or n-ary node is assigned to each child node unless the child node has a lower score. Control then continues to step S2246.

Please replace paragraph [0111] with the following amended paragraph:

[0111] In step S2510, any unique antecedent referents for the anaphors are determined. It will be apparent that anaphors such as pronouns usually follow their referent. Thus, any unique antecedent referent for the anaphor or pronoun usually occurs before the anaphor or pronoun in the structural representation of discourse. This reduces the number of preceding discourse constituents that must be examined. 20 Any determined unique antecedent referents are then substituted for the anaphor or pronouns. Control then continues to step S2525-S2515 and then immediately jumps to 25 step S2600 of Fig. 4.

Please replace paragraph [0119] with the following amended paragraph:

[0119] Fig. 14 is an exemplary structural representation of discourse for the 30 text to be summarized. The exemplary structural representation of discourse is comprised of conjoined discourse constituents. Line 10 of the exemplary text is

indicated as the important discourse constituent 1016910. Discourse constituents may be determined as important based on statistics, keywords, knowledge bases or any other non-structural measure of importance.

5 Please replace paragraph [0120] with the following amended paragraph:

[0120] Fig. 15 is an exemplary structural representation of discourse showing a first exemplary method of percolating relevance scores according to this invention. An important discourse constituent 1018 is identified based on some non-structural measure of relevance. The arrows indicate the percolation of the relevance score information indicated by an *R* value of "1" through a structural representation of discourse. Since the relevance score for the important discourse constituent 1018 is a child. ~~Relevance child, relevance~~ scores of more relevant child discourse constituents are assigned to the parent discourse constituent nodes. Since the important discourse constituent 1018 is the most relevant discourse constituent node, as determined by the selected non-structural measure of relevance, the relevance score of the important discourse constituent 1018 is assigned to the parent subordination node 1016. The relevance score of the parent node is in turn assigned to the coordination node 1012, the subordination node 1010, the coordination node 1002 and the root subordination node 10161000 based on the same rationale.

20

Please replace paragraph [0121] with the following amended paragraph:

[0121] Each of the subordinating discourse constituent nodes 1001, 1004, 1011, 1014 and 10161017 are assigned the relevance scores of the subordinated discourse constituent as indicated by the arrows. The relevance score of the most relevant child node is assigned to each of the coordination nodes. The coordination nodes 1002 and 1012 are included based on their status as parent nodes of the important discourse constituent. The relevance score is then percolated to the children of each of the identified coordination nodes 1002 and 1012. Thus, subordination nodes 1013 and 1011 and leaf nodes 1019 and 1020 are also assigned the relevance score. Each discourse constituent with a most relevant relevance score is selected for the hybrid text summary.

Please replace paragraph [0122] with the following amended paragraph:

[0122] Fig. 16 is an exemplary structural representation of discourse showing a second exemplary method of percolating relevance scores according to this invention. An important discourse constituent 1011016 is determined based on one or more non-structural measures of relevance. The percolation differs from the method described above in the treatment of the coordination and binary child nodes. More relevant relevance scores of the children of coordination or binary nodes are assigned to each preceding sibling. This reduces the number of included discourse constituents in the hybrid text summary 920.

10

Please replace paragraph [0126] with the following amended paragraph:

[0126] Fig. 20 is a first view of the exemplary structural representation of discourse showing percolation of relevance scores based on anaphor resolution according to this invention. Line 5 is identified as an important discourse constituent. In this method, relevance scores are restrictively percolated so that the adjustment of a discourse constituent relevance score occurs only if the discourse constituent can potentially resolve referents in the selected leaf discourse constituent node. Thus, line 4 is added since the term “a hot soup” potentially resolves the reference in line 5 to “the soup”. However, adding line 4, introduces the referent “they”. The term “they” is resolved by percolating the relevance score to line 3. Line 3 contains the term “Noodles”, which uniquely identifies the referent. In various other exemplary embodiments according to this invention, the term “Noodles” may be substituted into a line-line 4 to form a further compressed hybrid text summary 923.

25

Please replace paragraph [0127] with the following amended paragraph:

[0127] Each of the circuits 10-90 of the hybrid summarization system 100 systems 100-110 described in Figs. 3 and 5 can be implemented as portions of a suitably programmed general-purpose computer. Alternatively, 10-90 of the hybrid summarization system 100 systems 100-110 outlined above can be implemented as physically distinct hardware circuits within an ASIC, or using a FPGA, a PDL, a PLA or a PAL, or using discrete logic elements or discrete circuit elements. The particular form each of the circuits 10-90 of the hybrid summarization system 100 systems 100-

110 outlined above will take is a design choice and will be obvious and predictable to those skilled in the art.

Please replace paragraph [0128] with the following amended paragraph:

5 **[0128]** Moreover, the hybrid summarization ~~system 100-systems 100-110~~ and/or each of the various circuits discussed above can each be implemented as software routines, managers or objects executing on a programmed general purpose computer, a special purpose computer, a microprocessor or the like. In this case, the hybrid summarization ~~system 100systems 100-110~~ and/or each of the various circuits discussed above can each be implemented as one or more routines embedded in the communications network, as a resource residing on a server, or the like. The hybrid summarization ~~system 100-systems 100-110~~ and the various circuits discussed above can also be implemented by physically incorporating the hybrid summarization ~~system 100-systems 100-110~~ into software and/or a hardware system, such as the hardware and software systems of a web server or a client device.

10

15

Please replace paragraph [0130] with the following amended paragraph:

20 **[0130]** The communication links 99 shown in Figs. 1, 3 and 5 can each be any known or later developed device or system for connecting a communication device to the hybrid summarization ~~system 100systems 100-110~~, including a direct cable connection, a connection over a wide area network or a local area network, a connection over an intranet, a connection over the Internet, or a connection over any other distributed processing network or system. In general, the communication links 99 can be any known or later developed connection system or structure usable to connect devices and facilitate communication

25